LIVER

Development, validation, and evaluation of the PBC-40, a disease specific health related quality of life measure for primary biliary cirrhosis

A Jacoby, A Rannard, D Buck, N Bhala, J L Newton, O F W James, D E J Jones

See end of article for authors' affiliations

Correspondence to:
Professor A Jacoby, Division of Public Health, University of Liverpool, Whelan Building, The Quadrangle, Brownlow Hill, Liverpool L69 3GB, UK; ajacoby@liv.ac.uk

Background and aims: Study of health related quality of life (HRQOL) and the factors responsible for its impairment in primary biliary cirrhosis (PBC) has, to date, been limited. There is increasing need for a HRQOL questionnaire which is specific to PBC. The aim of this study was to develop, validate, and evaluate a patient based PBC specific HRQOL measure.

Subjects and methods: A pool of potential questions was derived from thematic analysis of indepth interviews carried out with 30 PBC patients selected to represent demographically the PBC patient population as a whole. This pool was systematically reduced, pretested, and cross validated with other HRQOL measures in national surveys involving a total of 900 PBC patients, to produce a quality of life profile measure, the PBC-40, consisting of 40 questions distributed across six domains. The PBC-40 was then evaluated in a blinded comparison with other HRQOL measures in a further cohort of 40 PBC patients.

Results: The six domains of PBC-40 relate to fatigue, emotional, social, and cognitive function, general symptoms, and itch. The highest mean domain score was seen for fatigue and the lowest for itch. The measure has been fully validated for use in PBC and shown to be scientifically sound. PBC patient satisfaction, measured in terms of the extent to which a questionnaire addresses the problems that they experience, was significantly higher for the PBC-40 than for other HRQOL measures.

Conclusion: The PBC-40 is a short easy to complete measure which is acceptable to PBC patients and has significantly greater relevance to their problems than other frequently used HRQOL measures. Its scientific soundness, shown in extensive testing, makes it a valuable instrument for future use in clinical and research settings.

Primary biliary cirrhosis (PBC) is a chronic cholestatic liver disease, which is progressive in a proportion of patients resulting in end stage liver disease. There is an increasing appreciation that the quality of life of PBC patients is frequently impaired by symptoms and other associated aspects of the disease. Typical symptoms include pruritus and fatigue which can occur at any point in the disease course and which may be independent of histological stage.[1–3] The effects of PBC on life quality have, however, been subjected to only limited formal analysis, not least because of the difficulties inherent in quantifying such subjective experiences. Where such formal studies have been performed they have tended to strongly confirm the view that this disease has an often dramatic negative effect on quality of life.[2 4–6] Reasonable conclusions that can be drawn from our current state of knowledge are, therefore, that quality of life is probably significantly impaired in a meaningful proportion of PBC patients and that appreciation of the significance of this quality of life impairment is critical if we are to fully understand their experiences. A clear implication of these conclusions is that study of the mechanisms responsible for quality of life impairment, and the development of treatments able to reverse these processes and improve quality of life, are important research goals in their own right. Critically, however, each of these issues (assessment of the impact of quality of life impairment on patients, identification of affected patients for studies of pathogenesis, and assessment of quality of life improving treatments in clinical trials) requires clinical tools which allow us to reliably, relevantly, and sensitively measure quality of life impairment in PBC patients.

The study of quality of life in PBC has conventionally been performed using existing health related quality of life (HRQOL) measures derived in other, often non-liver, chronic diseases.[2 4 7–12] The relevance and appropriateness of the application of these generic tools for the assessment of HRQOL to patients with PBC is, however, highly questionable, particularly since extrapolation of their use clinically has typically taken place with limited, if any, assessment of their validity in the context of PBC.[13] This is in addition to the general concern about the appropriateness of the application of generic HRQOL measures to patients with chronic diseases.[14–17] A further potential problem is that quality of life measures developed and tested in one or two treatment centres, or in a defined geographical area,[2–4 10 12] may have relevance to only those specific situations. These various limitations highlight the need for a robust PBC specific measure through which a wide picture of its impact on HRQOL can be documented. Such a measure could be used in studies of the pathways responsible for quality of life impairment and of their therapeutic modification. The aim of this study was therefore to derive, validate, and evaluate the use of a PBC specific HRQOL measure.

METHODS
Study design
The study had three phases (see fig 1). In the first phase, indepth interviews with patients were used to derive an initial measure, which was then reduced in size and refined following completion by a large patient cohort. In the second phase, the resulting measure (the PBC-40) was refined and validated in a further large patient survey. The aim of this phase of the study was to evaluate the scientific soundness of the new measure, as indicated by its psychometric properties (that is, validity, reliability, precision, and acceptability to patients—see below). In the third phase, the PBC-40 was evaluated in PBC patients in comparison with previously used HRQOL measures. The appropriateness of the patient orientated approach adopted is widely recognised.[18] [19] Ethics approval for the study was obtained from local research ethics committees in participating centres. Data protection requirements were observed throughout the study.

Participants
Participants taking part in indepth interviews and in the final evaluation stage were recruited from outpatient clinics in two secondary care centres in the North of England. All of these patients had definite or probable PBC defined using established diagnostic criteria (at least two of: (1) cholestatic liver function tests, (2) compatible or diagnostic liver histology, and (3) AMA at a titre of >1 in 80).[20] Participants for the two postal surveys conducted were recruited via the PBC Foundation, a national (UK-wide) charitable organisation providing support and information to people affected by PBC. Although it was not possible, for both ethical and logistic reasons, to verify their diagnostic status from clinical records, all postal survey participants were required to confirm that they had been diagnosed with PBC. A total of 985 PBC subjects participated in the different phases of the study.
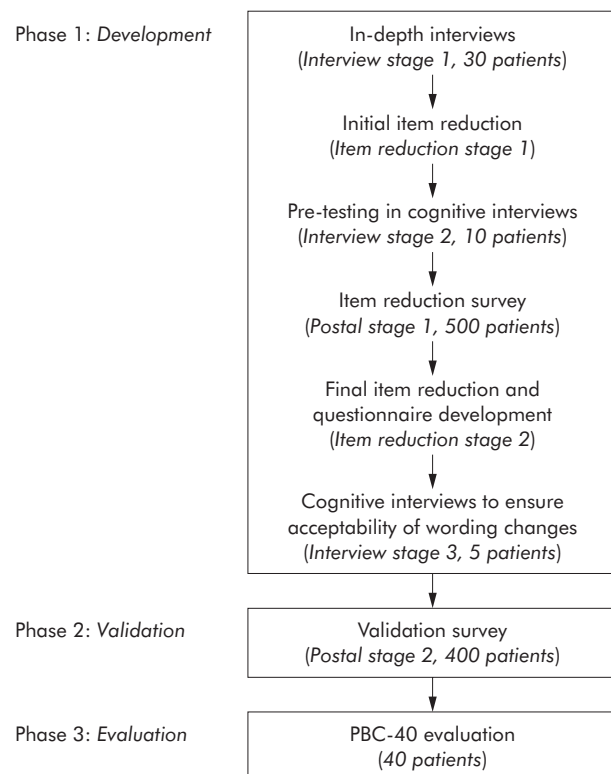
Phase 1: *Development*

In-depth interviews
(*Interview stage 1, 30 patients*)

Initial item reduction
(*Item reduction stage 1*)

Pre-testing in cognitive interviews
(*Interview stage 2, 10 patients*)

Item reduction survey
(*Postal stage 1, 500 patients*)

Final item reduction and
questionnaire development
(*Item reduction stage 2*)

Cognitive interviews to ensure
acceptability of wording changes
(*Interview stage 3, 5 patients*)

Phase 2: *Validation*

Validation survey
(*Postal stage 2, 400 patients*)

Phase 3: *Evaluation*

PBC-40 evaluation
(*40 patients*)

Figure 1 Flow diagram outlining the study design

Derivation phase
Potential component questions for the PBC-40 were generated from indepth interviews with 30 patients (interview stage 1), during which they were asked to describe the various ways in which having PBC affected them and their life quality. Following recognised sampling approaches for qualitative research[21] the 30 participants were selected for initial indepth interviews using age and time of diagnosis as sampling criteria. This cohort was therefore representative demographically of the local PBC patient population as a whole. Their accounts were thematically analysed to generate a series of potential questions, using the NVIVO 1.0 qualitative data analysis package.[22]

The complete list of potential questions derived from the indepth interviews then underwent a two step reduction process. The first step, which consisted of multiple rounds of editing performed by the authors, was used to discard questions that were repetitious, ambiguous, contained colloquial expressions or jargon, or which were deemed to reflect an individual's specific experience (question reduction stage 1). The questions remaining after this initial editing process were factored into related symptom/quality of life theme areas (domains). The questions were framed as a series of statements with prespecified response options (see table 1), from which respondents were instructed to select the option matching most closely their situation. Response options for all questions were on a standard five point Likert response scale[23] ranging from least burden or problem (score of 1) to greatest burden or problem (score of 5). A four week time frame was used for symptom specific questions, while responses to more general questions regarding social life and employment were not given a time frame.

The questions were first assessed in a further round of indepth "cognitive" interviews[24] in 10 patients to ensure ease of understanding (interview stage 2) and then distributed to 500 PBC patient members of the PBC Foundation (postal stage 1). The results of this survey were used to guide the second question reduction round (question reduction stage 2). Question reduction was performed using a previously described established method[25] (questions were retained at this stage if the frequency of endorsement for each response option was >5% and non-response for each question was <5%). To avoid redundancy, questions were removed if the correlation between any two was >0.8 (that is, they were functioning as duplicates). The result of this process was a draft HRQOL measure. A further final round of cognitive interviews (using five patients with definite PBC) was used to confirm the acceptability of the finalised wording of questions and their related response options (interview stage 3) before proceeding to the refinement and validation phases of the study.

Validation phase
The draft measure was validated against a battery of other widely used HRQOL and related measures and finalised. A further large group of PBC Foundation PBC patient members were asked to complete the full measure battery (including the PBC-40, the medical outcomes survey short form-36 (SF-36)[26] [27] previously used in the study of, among others, liver disease patients,[5–8] [10] [12] and three questions assessing cognitive function from another condition specific HRQOL measure (epilepsy surgery inventory-55 (ESI-55)[28]), together with 12 general questions regarding perception of their health status and quality of life (including impact of the disease on work) and information regarding their disease diagnosis (postal stage 2). A subgroup of patients participating in postal stage 2 were also asked to complete the measure on a second occasion two weeks after completing it for the first time to assess reproducibility. Where there were <50%

missing values for questions in each domain, we adopted a method advocated by McHorney and colleagues[25] for imputing mean values by adding the person specific mean for remaining question values within the domain. Where there were >50% missing values, domain responses were recorded as missing data.

The principal aim of the validation phase of the study was to evaluate the psychometric properties of the measure against established criteria.[13][29] The criteria specifically addressed were acceptability to patients, reliability (that is, that a measure is reproducible and internally consistent), validity (that is, that a measure measures what it purports to measure), and precision (that is, the ability of a measure to reflect true differences). Of the other recommended criteria,[13][29] appropriateness was inbuilt given that the questionnaire was developed exclusively in PBC patients; while responsiveness (that is, the ability of a measure to detect change over time) could not be tested as a result of the study design and the lack of any accepted therapies other than transplantation able to improve quality of life, against which the responsiveness could be appropriately assessed.

Acceptability was determined through pretesting the measure in cognitive interviews with PBC patients in terms of wording of questions and their related response options, and by the length of time taken to complete the measure. Response rates to the overall measure and to individual questions also provided an indication of the measure's acceptability.

Reliability was established through examining internal consistency or homogeneity of the questions in each domain using Cronbach's α coefficient*.[30] An α of at least 0.70 is considered necessary for a scale to be deemed internally consistent; although an α of 0.85 is considered optimal.[31] Correlations between individual question and total domain scores were examined as a further measure of internal consistency reliability.[32] It has been proposed by McHorney and colleagues[25] that a correlation >0.4 is acceptable for the purpose of establishing scale reliability. Test-retest reliability was calculated using intraclass correlation coefficients (ICCs). Coefficients of >0.9 are considered acceptable for individual comparisons while scores between 0.5 and 0.7 are adequate for group comparisons.[25][33] We used a two week interval between assessments, which is generally considered to be a suitable time frame.[34]

Validity was examined in three respects, relating to (a) face validity (ensuring that the instrument was measuring what it was supposed to measure), (b) content validity (ensuring that the instrument contained sufficient questions to adequately assess HRQOL in PBC), and (c) construct validity (ensuring that the measure correlated in a predicted way with other measures in the test battery). Face and content validity were established by ensuring that the patient generated questions contained in the PBC-40 addressed dimensions of HRQOL in PBC that were important and relevant to patients themselves. Construct validity was established by examining correlations of the PBC-40 with the cognition questions from two other well recognised HRQOL measures (SF-36 and ESI-55), using Pearson correlation coefficients. We predicted a low to moderate correlation between the social domain of the PBC-40 (10 questions) and the social function domain of the SF-36 (two questions) due to differences in the number of questions

....

*Cronbach's coefficient *alpha* is a measure of the relatedness of items (questions) in a multi-item scale. High *alpha* coefficients indicate high relatedness (that is, internal consistency) of the items. To examine their contribution to the scale in question, items are systematically removed one at a time. If their removal leads to an increase in the *alpha* coefficient (that is, improved internal consistency), items are then discarded.

between the two scales and the contrast in recall parameters for the PBC-40 (general) and the SF-36 (responses limited to the last four weeks). We predicted moderate to high correlations between the fatigue domain of the PBC-40 and the vitality/energy domain of the SF-36; a moderate to high correlation between the emotional domain of the PBC-40 and the role emotional and mental health domains of the SF-36; a high correlation between the cognitive domain of the PBC-40 and the cognition questions of the ESI-55; and a moderate to high correlation between the symptoms and itch domains of the PBC-40 and the physical domains of the SF-36. Associations between the PBC-40 domain scores and the 12 general questions relating to global quality of life, employment, and health status were also examined.

Precision of the measure was established by examining score distributions for floor or ceiling effects. A floor effect occurs where respondents are frequently scoring at the bottom of the scale (low degree of HRQOL impairment) with the effect that subsequent improvement (that is, following therapeutic intervention) cannot be measured. A ceiling effect represents the converse situation where respondents are scoring at the top of the scale (high degree of HRQOL impairment) with the effect that subsequent deterioration in HRQOL cannot be detected.

Evaluation phase

In order to evaluate the scope and relevance of the PBC-40 as a clinical tool in comparison with other quality of life and symptom assessment tools previously used in PBC, a further group of 40 patients (all with definite PBC) who had not participated in its derivation was given anonymised versions of the SF-36,[26][27] PBC-40, fatigue impact scale (FIS),[9] and chronic liver disease questionnaire (CLDQ)[10] to complete. The measures were arranged in random sequence to avoid potential questionnaire order effects. Participants were asked to score each measure in response to the question "How well does this questionnaire address the problems you have encountered as a result of having PBC?" Responses were on an integral scale of 1 to 10 with the label 1 representing "not at all" and 10 representing "very well". Individual satisfaction scores for each measure were compared using one way ANOVA. Frequencies with which respondents identified each measure as that which most accurately addressed their problems, and with which respondents identified a measure as addressing their problems poorly (defined as a response score <5) were compared by χ^2 test.

RESULTS
Derivation and validation of the PBC-40

The 30 initial indepth interviews (interview stage 1) generated 2498 respondent statements of potential use as HRQOL measure questions. A total of 2317 of these potential questions were discarded in the initial reduction process (question reduction stage 1). The remaining 180 questions were pretested in cognitive interviews with 10 patients (interview stage 2). Thirty nine questions were abandoned after these interviews because patients felt they were ambiguous, unclear, or difficult to answer. The remaining 141 questions were then tested in a national postal survey (postal stage 1).

Of the 500 questionnaires sent out to PBC Foundation members self-reporting a diagnosis of PBC, 378 (75%) were returned. Of these 336 were usable in the study (the other 42 patients had either received liver transplants, had died (and the questionnaires were returned blank by relatives), or had already been involved in earlier phases of the study). Mean age of respondents in this postal survey was 59 (10) years with 312 (93%) females, 20 (5%) males, and no gender data for four patients. Following this survey, 84/141 questions

Table 1 The six domains and 40 questions of the PBC-40†

Domain	Question	Response options*
Symptoms	*In the last 4 weeks:* I was able to eat what I liked I ate or drank only a small amount, and still felt bloated I felt unwell when I drank alcohol I had discomfort in my right side I had dry eyes My mouth was very dry I had aches in the long bones of my arms and legs	Never, rarely, sometimes, most of the time, always
Itch	*In the last 4 weeks:* Itching disturbed my sleep I scratched so much I made my skin raw I felt embarrassed because of the itching	Never, rarely, sometimes, most of the time, always
Fatigue	*In the last 4 weeks:* I had to force myself to get out of bed I had to have a sleep during the day Fatigue interfered with my daily routine I felt worn out I felt so tired, I had to force myself to do the things I needed to do I felt so tired, I had to go to bed earlier than usual Fatigue just suddenly hit me PBC drained every ounce of energy out of me Some days it took me a long time to do anything If I was busy one day I needed at least another day to recover I had to pace myself for day-to-day things	Never, rarely, sometimes, most of the time, always
Cognition	*In the last 4 weeks:* I had to make a lot of effort to remember things I had difficulty remembering things from one day to the next My concentration span was short because of PBC I had difficulty keeping up with conversations I found it difficult to concentrate on anything I found it difficult to remember what I wanted to do	Never, rarely, sometimes, most of the time, always
Social	My sex life has been affected by having PBC I feel I neglect my family because of having PBC I feel guilty that I can't do what I used to do because of having PBC I sometimes feel frustrated that I can't go out and enjoy myself I tend to keep the fact that I have PBC to myself I can't plan holidays because of having PBC My social life has almost stopped Everything in my life is affected by PBC PBC has reduced the quality of my life I can still lead a normal life, despite having PBC	Not at all, a little, somewhat, quite a bit, very much
Emotional	Because of PBC, I get more stressed about things than I used to Having PBC gets me down I worry about how my PBC will be in the future	Not at all, a little, somewhat, quite a bit, very much

*Items are scored from 1 to 5 and the individual item scores are summed to give a total domain score. Maximum and minimum domain scores are shown in table 5. The direction of scoring of some items is reversed for calculation of domain scores so that in all cases, high scores represent high impact and low scores low impact of PBC on quality of life.
†Copies of the questionnaire and full scoring instructions are obtainable from the authors.

were abandoned through principal components analysis (question reduction stage 2). Following comments returned in the postal survey, appropriate alterations were also made to the wording of some of the remaining questions and to the format of response options (for example, agree/disagree options were replaced with never/always options, which respondents found to be a more appropriate form of response). A second round of cognitive interviews with five patients tested the reworded questions for acceptability (interview stage 3). The remaining 57 questions were then combined with the SF-36, three cognition questions from the ESI-55, and the 12 general questions to form a test battery for validation in a second national postal survey (postal stage 2).

In this second postal survey, 260/400 patients (65%) returned survey questionnaires. Of these 240 were usable. Mean age of respondents was 60 (10) years with 223 (92%) females, 12 (5%) males, and no gender data for five patients. Of the 400 subjects invited to participate in the second postal survey, 100 were asked to complete a further questionnaire two weeks later to allow test/retest reliability testing. Of these 100 test/retest subjects, 89 returned questionnaires, all of which were useable. Mean age of the retest group was 62 (10) years with 94% female and 6% male. Based on responses to the second postal survey, three questions relating to employment were abandoned because the level of overall endorsement for each was <5%. Principal component analysis performed on the remaining 54 questions identified a 10 factor structure from which a further 16 were discarded because they loaded on two or more factors. Two of these questions (''I had aches in the long bones of my arms and legs'' and ''Having PBC gets me down'') were later reintroduced on the grounds of content validity. The final version of the measure thus consisted of 40 questions, each scored on a scale of 1 to 5 (where 1 = least impact, 5 = greatest impact) grouped into six domains (symptoms, itch, fatigue, cognition, social, and emotional) (table 1). For

Table 2 Reliability of the PBC-40

Domain	Cronbach's α	Corrected item-total correlations (range)	Test-retest (ICC)
Social	0.89	0.19–0.76	0.96
Fatigue	0.94	0.58–0.86	0.96
Cognitive	0.95	0.82–0.88	0.95
Emotion	0.74	0.50–0.70	0.83
Itch	0.94	0.86–0.90	0.92
Symptoms	0.72	0.22–0.53	0.95

Internal consistency was established using Cronbach's α coefficient[30] (a minimum of 0.7 is recommended for group comparisons and a minimum of 0.9 for individual comparisons)[31] and corrected item-total correlations[32] (a correlation of >0.4 is acceptable for the purpose of establishing scale reliability).[25] Test-retest reliability was calculated using intraclass correlation coefficients (ICCs). Coefficients of >0.9 are considered acceptable for individual comparisons while scores between 0.5 and 0.7 are adequate for group comparisons.[25 33] A standard two week interval between assessments was used.[34]

Table 3 Validity of the PBC-40

Comparator measure domain	PBC-40 domain	Pearson correlation coefficient
ESI-55 Cognition	Cognitive	−0.81*
SF-36 Social functioning	Social	−0.80*
SF-36 Physical functioning	Symptoms	−0.57*
SF-36 Physical functioning	Itch	−0.21
SF-36 Energy/vitality	Fatigue	−0.76*
SF-36 Mental health	Emotional	−0.58*
SF-36 Role emotional	Emotional	−0.50*

Validity was determined by the degree of correlation between the domains of the PBC-40 and key relevant domains of the short form-36 (SF-36) and the epilepsy surgery inventory-55 (ESI-55).
*Correlation significant at the 0.01 level (two tailed).

each domain, scoring involved summing individual question response scores (the number of questions in each domain and the possible domain score ranges are shown in table 4); with higher scores indicating poorer quality of life.

Assessment of the psychometric properties of the PBC-40

The potential usefulness of any HRQOL measure is determined by its psychometric properties.[13] We therefore went on to assess the psychometric properties of the finalised PBC-40 using the data from postal stage 2 (260 patients, 89 with test-retest completion) and from the cognitive interviews.

Appropriateness

Developing the measure solely with a PBC population ensured its appropriateness and relevance for this patient group. Questions were based on what patients said in indepth interviews, and further testing in cognitive interviews ensured that both questions and response options were appropriate for the final measure.

Acceptability

The average 10–20 minute completion time of the PBC-40 in cognitive interviews was found to be acceptable to respondents. Overall response rates were >65% for each phase of the study, a rate considered acceptable for postal surveys.[35] The validation survey showed 45/240 cases with missing values, and of these only four had missing values >50% (where it appeared that respondents had turned over two pages of the questionnaire together). The low range of missing values for the test battery (PBC-40 0.4%–4.2%; SF-36 0.4%–2.9%; ESI-55 0.8%) also indicated good acceptability to respondents.

Reliability

Cronbach's α scores in this sample ranged from 0.72 to 0.95 and so exceeded the standard recommended for group comparison on all six domains (table 2). Three of the domains (fatigue, cognitive, and itch) met the recommended minimum of 0.9 for individual patient comparisons. ICC scores ranged from 0.83 to 0.96. Question total correlations were acceptable in all but two domains. In the social domain, the correlation for the question "I tend to keep the fact that I have PBC to myself" was only 0.19, while the symptoms domain question "I felt unwell when I drank alcohol" had a correlation of only 0.22. The low correlations for these two particular questions reflects the fact that they had originally been identified in factors 7–10 of the principal components analysis but had been placed with other similar questions in the social and other symptoms domains for reasons of face validity.

Validity

Developing and pretesting the PBC-40 with patients themselves established face and content validity. Construct validity of the PBC-40 was demonstrated by the high correlations seen with the SF-36 and ESI-55 cognition questions (table 3). Correlation between the social domain of the PBC-40 and the social function domain of the SF-36 was higher than anticipated given the differences in question number between the scales and the different recall parameters for the PBC-40 and SF-36. The predicted moderate to high correlation between the fatigue domain of the PBC-40 and the vitality/energy domain of the SF-36 was confirmed. Similarly, the moderate to high correlations between the emotional domain of the PBC-40 and the role emotional and mental health domains of the SF-36 were as predicted. High correlation between the cognitive domain of the PBC-40 and the cognition questions of the ESI-55 was also as predicted.

Precision

There were no significant ceiling effects but a moderate floor effect was found for the itch domain (36.7%). The extent of floor and ceiling effects is shown in table 4, together with the range of actual domain scores, mean domain scores, and standard deviations.

Table 4 Precision of the PBC-40

Domain (No of questions)	Possible range	Actual range	Mean score	SD	Floor effect (%)	Ceiling effect (%)
Fatigue (11)	11−55	11−54	34.6	9.5	1.3	0.0
Social (10)	10−50	11−49	28.5	9.2	0.0	0.0
Symptoms (7)	7−35	7−31	18.5	5.05	2.5	0.0
Cognitive (6)	6−30	6−30	15.7	6.15	12.5	2.9
Emotion (3)	3−15	3−15	9.08	3.2	2.0	5.4
Itch (3)	3−15	3−15	6.02	3.32	36.7	1.7

Estimates of floor and ceiling effects (numbers of respondents with the minimum possible and maximum possible scores for each domain, respectively).
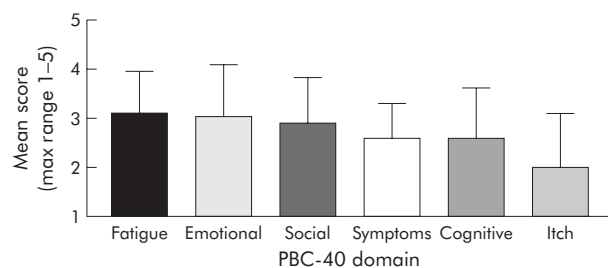
Figure 2 Mean scores for each domain of the PBC-40 in 260 patients. Error bars denote SD. In this cohort the fatigue domain had the highest mean score and the itch domain the lowest mean domain score.

In the 260 PBC patients participating in postal stage 2, no significant correlations were seen between any PBC-40 domain score and age, self reported disease stage, length of time since diagnosis, or sex (data not shown). The general question that asked "Did you have to give up work because of PBC" was, however, significantly associated with all domain scores ($p < 0.001$ for all domains; data not shown). In these 260 patients the fatigue domain was found to have the highest mean score and the itch domain the lowest (fig 2).

Evaluation of the PBC-40

Of the 40 patients participating in the evaluation phase of the study, 35 returned fully completed questionnaires (88%, mean age 62 (11) years; 32 females, three males). Mean satisfaction score was significantly higher for the PBC-40 than for the other measures ($p < 0.0005$, one way ANOVA). The PBC 40 was ranked highest (or equal highest) of the measures with regard to the extent to which it addressed their problems by significantly more of the patients (83%) than any of the other measures (fig 3A). Conversely, the percentage of patients rating the PBC-40 as unsatisfactory (defined as a satisfactions score of less than 5) was significantly lower (3%) than for any of the other measures (fig 3B). No relationship was seen between patient age and degree of satisfaction for any of the measures.

DISCUSSION

We have developed a brief psychometrically robust measure of HRQOL, the PBC-40, intended specifically for use with patients with PBC. The measure was developed using currently recommended patient centred methods, and is intended for use in both clinical and research settings. We believe that it will prove a valuable addition to the range of such measures on which those working with PBC patients can draw. The PBC-40 is a profile measure, covering six PBC specific quality of life domains (cognitive, social, emotional function, fatigue, itch, and other symptoms). Future work will explore the possibility of summing the individual

domains in order to produce an overall summary score for quality of life. This approach has recently been proposed for the widely used generic HRQOL measure the SF-36.[36]

It is now generally accepted that the effects of a chronic illness such as PBC and of its treatment should be measured in terms of quality not just quantity of survival, and that failure to do so is neither "good science nor good medicine".[37] Formal assessment of quality of life in PBC in both clinical and research settings requires access, however, to measurement tools that are relevant to the disease, scientifically sound, and patient centred enough to make such assessments meaningful. The fundamental importance of the role played by patients in the development process if the last of these aims is to be achieved is now widely recognised and has guided the approach adopted in this study.[18 38 39] In the current study, we ensured that our PBC specific quality of life measure had high validity by employing what have been referred to as discovery methods[19] and involving patients directly in the identification of quality of life domains to be covered. We did so both at the stage of initial question generation by using indepth interviews and during the subsequent question reduction and validation stages, by using cognitive interviewing techniques. This approach, which allowed patients to say what was important and what was unimportant in our proposed question set, what was missing, and what was present but redundant, has previously been identified as being of significant value in the development of high quality patient centred quality of life measures.[40 41]

Following completion of the derivation steps, the PBC-40 was validated in a large patient cohort and its psychometric properties assessed. These studies show the PBC-40 to be highly psychometrically robust. The extent of these validation studies (involving almost 1000 PBC patients (approximately 5% of the estimated total UK PBC patient population, making this the largest single clinical study performed in this condition to date) make the PBC-40 the HRQOL measure with the best validation for use in PBC to date. Six psychometric properties of HRQOL measures have been identified as being importance in determining the utility and value of such measures.[13]

The appropriateness of the PBC-40 was ensured by performing all development steps exclusively in PBC patients. The acceptability of the PBC-40 was demonstrated by the feedback at cognitive interviews, and the return rate of >65% for all postal completion phases of the study. The reliability of the PBC-40 was demonstrated by Cronbach's α scores exceeding the recognised significance threshold of 0.7 for group comparisons for all six domains[31] (with fatigue cognitive and itch domains exceeding 0.9). ICC scores ranged from 0.83 to 0.96, demonstrating high test/retest reliability, again significantly exceeding recognised thresholds.[25] The validity of the PBC-40 was demonstrated by high correlations
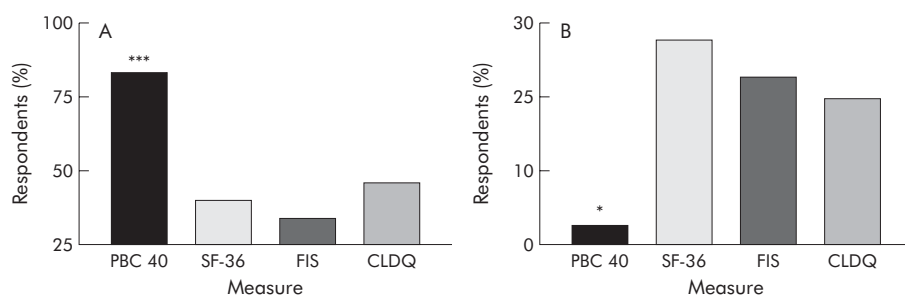

Figure 3 Percentages of primary biliary cirrhosis (PBC) patients who (A) ranked each measure highest or equal highest in terms of the degree to which it addressed their problems and (B) identified each measure as unsatisfactory in terms of the degree to which it addressed their problems. *$p < 0.05$, ***$p < 0.0005$.

with relevant domains of the SF-36 and ESI-55. The precision of the PBC-40 was demonstrated by the low levels of floor and ceiling effects. The precision of an HRQOL measure is a key issue if it is to be of use as a clinical trial outcome measure or in studies of natural history of the disease. Significant floor effects (the presence of the minimal possible scores in patients at study outset) preclude the measurement of any improvement in patients while significant ceiling effects (the converse presence of maximum scores at study outset) preclude the measurement of any deterioration in symptoms. Ceiling effects were almost entirely absent (an average of 1.7% of the domain scores from the 260 patients participating in the second postal stage were at the maximum level, with maximum score absent from all participants for the fatigue, social, and symptom domains). This suggests that the PBC-40 is likely to be of particular use as a clinical trial outcome measure. Floor effects were seen more frequently and were particularly marked in the case of the itch domain (37% of patients registering the lowest possible score). We believe that this effect is likely to reflect the rather dichotomous nature of itch in PBC, and the availability and efficacy of treatments for cholestatic itch (in contradistinction to the other less easily treatable symptoms of PBC) rather than a weakness of the PBC-40.

The final psychometric property of HRQOL measures, that of responsiveness, could not be tested in the PBC-40. Responsiveness can only be assessed by studying the degree to which a measure can detect and quantify a beneficial response to an established treatment previously proven to be able to improve outcome in the disease of interest. Unfortunately, no such treatments exist for the symptoms of PBC[42] with the exception of transplantation, a prospective trial of which was beyond the scope for this study. Indeed, the imperative to be able to develop and test such treatments in the future was one of the driving forces for the current study.

Following validation of the PBC-40 we set out to evaluate it in comparison with other HRQOL measures previously used in PBC (FIS, SF-36, and CLDQ). Participants in this phase were blinded with regard to the specific issue being addressed and were simply asked to score anonymised and randomly ordered versions of the four measures on a scale of 1 to 10 regarding the extent to which each measure addressed the problems relevant to them (they were not at this stage asked to complete the measures). The PBC-40 was identified as the most relevant measure by over 80% of patients. Only 3% of patients felt that the PBC-40 did not address their problems. In each regard the PBC-40 scored significantly more favourably than all of the other measures. Several factors are likely to underpin the greater patient satisfaction seen for the PBC-40 than for the other PBC applied measures. Unlike the PBC-40, the other measures studied had not explored patient views of symptoms in any depth and had been developed in secondary and tertiary care centres that were likely to represent a biased sample of individuals with PBC. Although the component questions for the PBC-40 were initially generated with patients drawn from hospital out-patient databases, results of subsequent testing in two national samples supported the fact that questions in the measure were meaningful and relevant to a wider UK audience (further validation studies will have to be performed before the PBC-40 can appropriately be used in non-UK populations). Our rigorous approach to question generation, reduction, and testing ensured that the measure also met the other currently specified criteria for quality of life measure development and application[29] In particular, we would draw attention to our use of indepth interviews with a purposively sampled group of PBC patients as the source of potential questions, and of adoption of cognitive interviewing

techniques to identify problems with wording of both the proposed questions and the response options.[24 43 44] We would suggest that both these approaches contributed to the outcome that patients found the PBC-40 relevant to their experiences acceptable and easy to complete, as evidenced by the low overall missing value scores in both the question reduction and validation surveys.

There are some limitations to the study. The high numbers of patients who had retired or who had given up work as a result of PBC meant that three questions relating to employment were abandoned due to low response rates. The PBC-40 may therefore not fully address the HRQOL of those that struggle to remain in employment despite the effects of PBC. However, the significant effect found between the general question related to giving up work because of PBC and all domains of the PBC-40 indicates that this question could be used as an indicator of discriminant validity. Many of our respondents commented on their inability to separate the effects of PBC from other conditions. For some, however, other conditions had preceded PBC for many years and they were able to compare current quality of life with what life was like before developing PBC. The effect of comorbidity is something that needs to be addressed further in future research.

In conclusion, we suggest that the PBC-40 represents a potentially important addition to already available HRQOL measures, being the only one that is truly PBC specific. We hope its use in clinical and research settings will allow for a more meaningful assessment of the quality of life of those affected by this distressing and often poorly understood condition. We believe that the PBC-40 meets the essential requirements to become a standard tool for quality of life outcome assessment in future clinical trials in PBC.

. . . .
Authors' affiliations
A Jacoby*, Division of Public Health, University of Liverpool, Liverpool, UK
A Rannard*, School of Nursing, Midwifery and Health Visiting, University of Manchester, Manchester, UK
D Buck, Division of Primary Care, University of Liverpool, Liverpool, UK
N Bhala, J L Newton, O F W James, D E J Jones, Liver Research Group, School of Clinical Medical Sciences, University of Newcastle, Newcastle, UK

*A Jacoby and A Rannard contributed equally to this study.

Conflict of interest: None declared.

REFERENCES
1 **Jones EA**, Bergasa NV. The pruritus of cholestasis. *Hepatology* 1999;**29**:1007–12.
2 **Cauch-Dudek K**, Abbey S, Stewart DE, *et al.* Fatigue in primary biliary cirrhosis. *Gut* 1998;**43**:705–10.
3 **Goldblatt J**, Taylor PJS, Lipman T, *et al.* The true impact of fatigue in primary biliary cirrhosis: a population study. *Gastroenterology* 2002;**122**:1235–41.
4 **Huet P**, Deslauriers J, Tran A, *et al.* Impact of fatigue on the quality of life of patients with primary biliary cirrhosis. *Am J Gastroenterol* 2000;**95**:760–7.
5 **Younossi ZM**, Kiwi ML, Boparai N, *et al.* Cholestatic liver diseases and health-related quality of life. *Am J Gastroenterol* 2000;**95**:497–502.

6 **Poupon RE**, Chretien Y, Chazouilleres O, *et al.* Quality of life in patients with primary biliary cirrhosis. *Hepatology* 2004;**40**:489–94.

7 **Unal G**, deBoer J, Borsboom G, *et al.* A psychometric comparison of health-related quality of life measures in chronic liver disease. *J Clin Epidemiol* 2001;**54**:587–96.

8 **Gralnek I**, Hays R, Kilbourne A, *et al.* Development and evaluation of the liver disease quality of life instrument in persons with advanced, chronic liver disease. *Am J Gastroenterol* 2000;**95**:3552–65.

9 **Prince M**, James O, Holland N, *et al.* Validation of a fatigue impact score in primary biliary cirrhosis: towards a standard for clinical and trial use. *J Hepatol* 2000;**32**:368–73.

10 **Younossi Z**, Guyatt G, Kiwi M, *et al.* Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease. *Gut* 1999;**45**:295–300.

11 **Kim WR**, Lindor K, Malinchoc M, *et al.* Reliability and validity of the NIDDK-QA instrument in the assessment of quality of life in ambulatory patients with cholestatic liver disease. *Hepatology* 2000;**32**:924–9.

12 **Gross CR**, Malinchoc M, Kim WR, *et al.* Quality of life before and after liver transplantation for cholestatic liver disease. *Hepatology* 1999;**29**:356–64.

13 **Rannard A**, Buck D, Jones DEJ, *et al.* Assessing quality of life in primary biliary cirrhosis. *Clin Gastroenterol Hepatol* 2004;**2**:164–74.

14 **Patrick DL**, Deyo RA. Generic versus disease-specific measures in assessing health status and quality of life. *Medical Care* 1989;**27**:S217–32.

15 **Avis NE**, Smith KW. Conceptual and methodological issues in selecting and developing quality of life measures. *Adv Med Sociol* 1994;**5**:255–80.

16 **Guyatt GH**, King DR, Feeny DH, *et al.* Generic and specific measurement of health-related QOL in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999;**52**:187–92.

17 **Kersten P**, Mullee MA, Smith JAE, *et al.* Generic health status measures are unsuitable for measuring health status in severely disabled people. *Clin Rehabil* 1999;**13**:219–28.

18 **Carr AJ**, Higginson IJ. Are quality of life measures patient-centred? *BMJ* 2001;**322**:1357–60.

19 **Kessler RC**, Mroczek DK. Some methodological issues in the development of quality of life measures for the evaluation of medical interventions. *J Eval Clin Pract* 1996;**2**:181–91.

20 **James OFW**, Bhopal R, Howel D, *et al.* Primary biliary cirrhosis once rare, now common in the UK? *Hepatology* 1999;**30**:390–4.

21 **Murphy E**, Dingwall R, Greatbach D, *et al.* Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998;**2**:16.

22 **Richards L**. *Using NVIVO in qualitative research.* Thousand Oaks: Sage Publications, 1999.

23 **Likert R**. A technique for the measurement of attitudes. *Arch Psychol* 1932;**140**:1–55.

24 **Harris-Kojetin LD**, Fowler FJ, Brown JA, *et al.* The use of cognitive testing to develop and evaluate CAHPS™ 1.0 Core Survey items. *Med Care* 1999;**37**(Suppl 3):MS10–21.

25 **McHorney CA**, Ware JE Jr, Lu JF, *et al.* The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;**32**:40–66.

26 **Ware JE**, Sherbourne CD. The MOS 36-item short form health survey (SF-36); 1: conceptual framework and item selection. *Med Care* 1992;**30**:473–83.

27 **Jenkinson C**, Layte R, Wright L, *et al.* *The UK SF-36: An analysis and interpretation manual.* Oxford: University of Oxford, 1996.

28 **Vickrey BG**. A procedure for developing a quality-of-life measure for epilepsy surgery patients. *Epilepsia* 1993;**34**(Suppl 4):S22–7.

29 **Aaronson NK**, Alonso J, Burnam MA, *et al.* Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qua Life Res* 2002;**11**:193–205.

30 **Cronbach LJ**. Coefficient alpha and the internal structure of tests. *Psychometrika* 1954;**16**:297–334.

31 **Nunnally J**, Bernstein JC. *Psychometric theory*, 3rd Edn., New York: McGraw-Hill 1994.

32 **Juniper ER**, Guyatt GH, Jaescke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, ed. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven, 1996.

33 **Stewart AL**, Hays RD, Ware JE. Methods of constructing health measures. In: Stewart AL, Ware JE, eds. *Measuring functioning and well-being. The medical outcomes study approach*. London: Duke University Press, 1992:49–56.

34 **Streiner DL**, Norman GR. *Health Measurement Scales*. 2nd Edn, Oxford: Oxford University Press, 1995.

35 **Mangione TW**. *Mail surveys—improving the quality, Applied Social Research Methods Series, vol.40*. Thousand Oaks: Sage Publications, 1995.

36 **Simon GE**, Revicki DA, Grothaus L, *et al.* SF-36 summary scores: are physical and mental health truly distinct? *Med Care*, **36**:445–8.

37 **Fallowfield L**. *The quality of life: the missing measurement in health care*. London: Souvenir Press, 1990.

38 **Entwistle VA**, Renfrew MJ, Yearley S, *et al.* Lay perspectives: advantages for health research. *BMJ* 1998;**316**:463–96.

39 **Goodacre H**, Smith R. Rights of the patient in research. *BMJ* 1995;**310**:1277–8.

40 **Guyatt GH**, Cook DJ. Health status, quality of life and the individual. *JAMA* 1994;**272**:630–1.

41 **Trauer T**, Mackinnon A. Why are we weighting? The role of importance ratings in quality of life measurement. *QOL Research* 2001;**10**:579–85.

42 **Jones DEJ**. Complications of cholestasis. *Medicine* 2002;**30**:67–8.

43 **Meadows KA**. Assessing health-related quality of life—what can the cognitive sciences contribute? An advanced seminar. *Qual Life Newsletter* 2001;**6**:2.

44 **Buck D**, Jacoby A, Massey A, *et al.* Pre-testing items in a stroke-specific quality of life measure: the value of cognitive interviewing. *J Neurolinguistics* 2000;**13**:267–71.

EDITOR'S QUIZ: GI SNAPSHOT ....

Answer

From question on page 1596

Computed tomography of the abdomen revealed a well defined rounded area of fluid collection anterior to the pancreatic uncus and two rounded calcified masses in the duodenal region (fig 2). Endoscopic sonography demonstrated an anechoic double walled cystic lesion adjacent to the pancreas. Based on radiological findings, a diagnosis of duodenal duplication cysts was considered and exploratory surgery was performed. At surgery, separation of the mass from the pancreas was difficult. Histopathological examination confirmed duodenal duplication cyst and found two other cysts containing enteroliths. Enteroliths were previously described in the natural history of duodenal duplication in adults. Intestinal duplication cysts are uncommon congenital anomalies and usually presents during the firsts decade of life. The majority of duplication cysts are tubular or spherical without any communication with the gastrointestinal tract. Duplication cysts of the duodenum are very rare, accounting for only 4–5% of intestinal duplications. Endoscopic sonography is useful, demonstrating anechoic mass surrounded by a double layer intestinal wall.

doi: 10.1136/gut.2005.068502